# Costumer Churn Prediction.

. Ade Dwi Aryani, batch 008

CHU

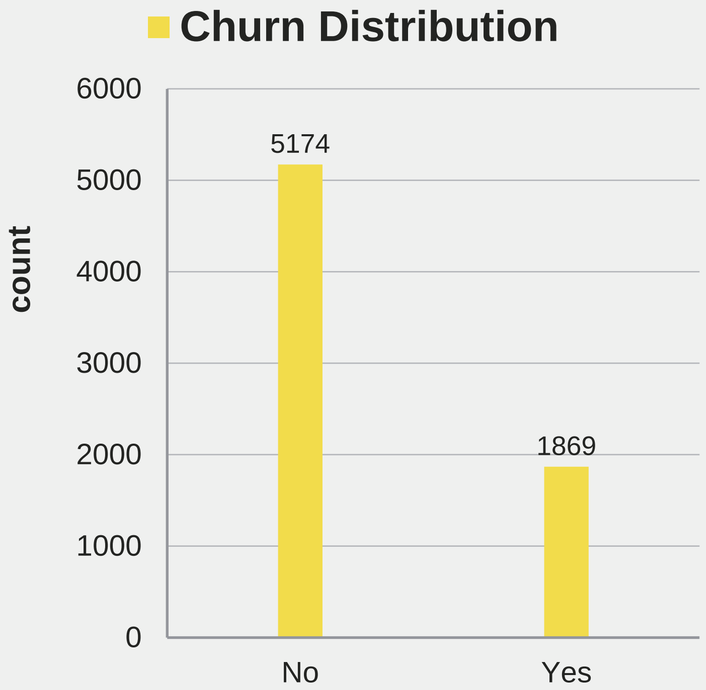
RN

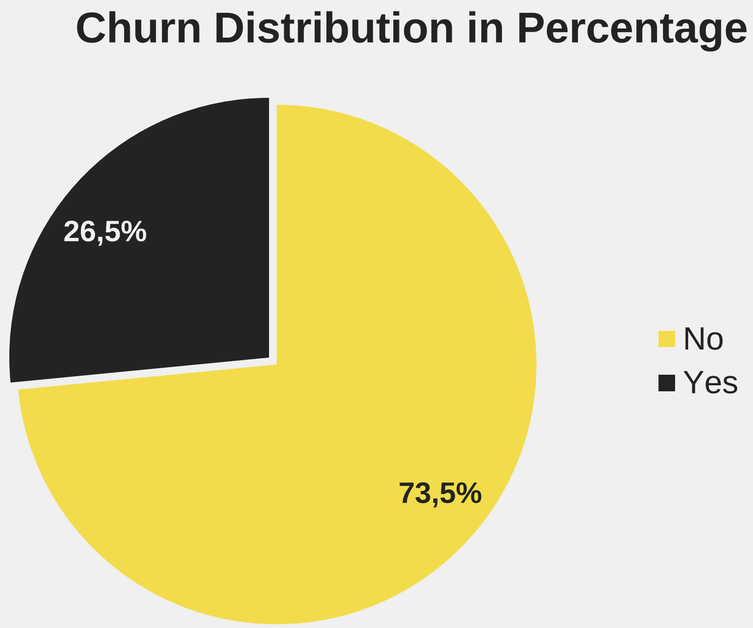
# Introduction

**Objective** : to predict customer churn with several relevant customer data.
The data to be used is **Telco Customer Churn Dataset**.
**The data set includes information about**:

- Customers who left within the last month – the column is called Churn
- **Basic information** about customers – gender, senior citizen, and if they have partners and dependents
- **Services** that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
- **Customer account** information – how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges

# Check for imbalance churn distribution

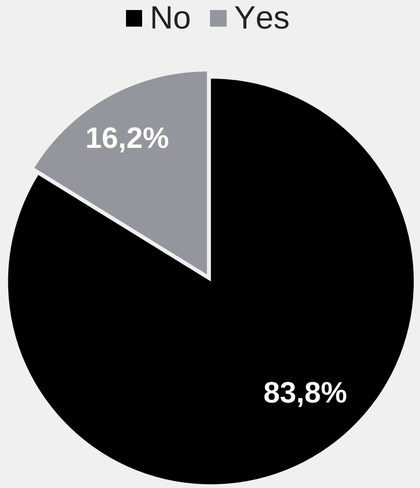



- The target variable has 73,5 % instances of ‘no-churn’.
- The target variable is imbalanced.

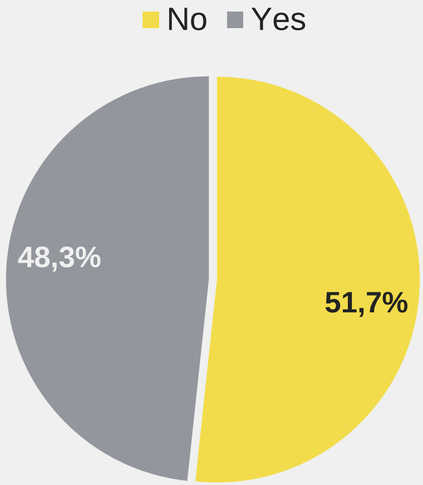
# Basic Information

## Customer by Senior Citizen



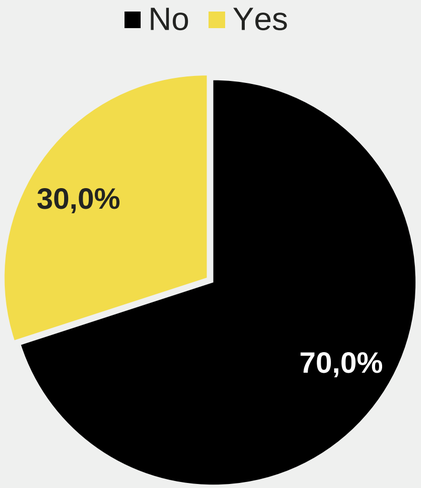

- There are only 16,2% of the customers who are senior citizens. Thus most of customers in dataset are younger people.

## Customer by Partner



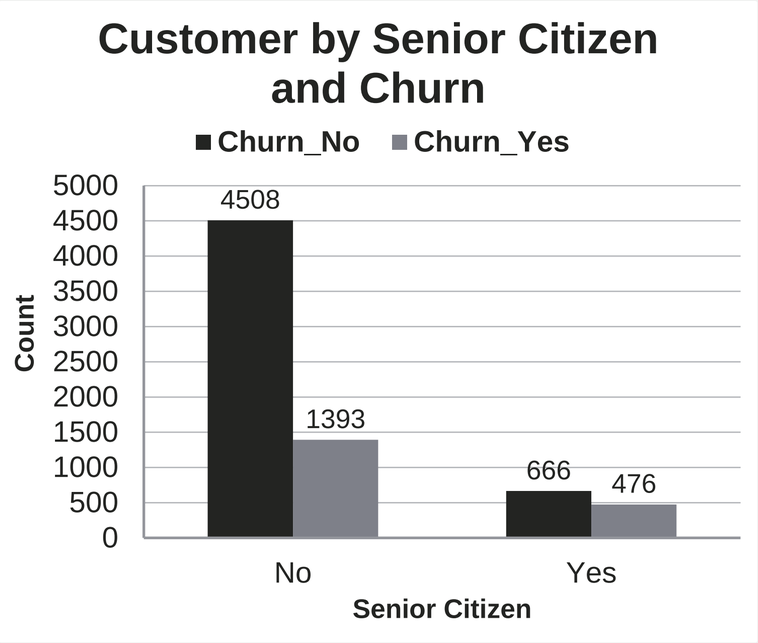

- About 48,3% of the customers have a partner.

## Customer by Dependents

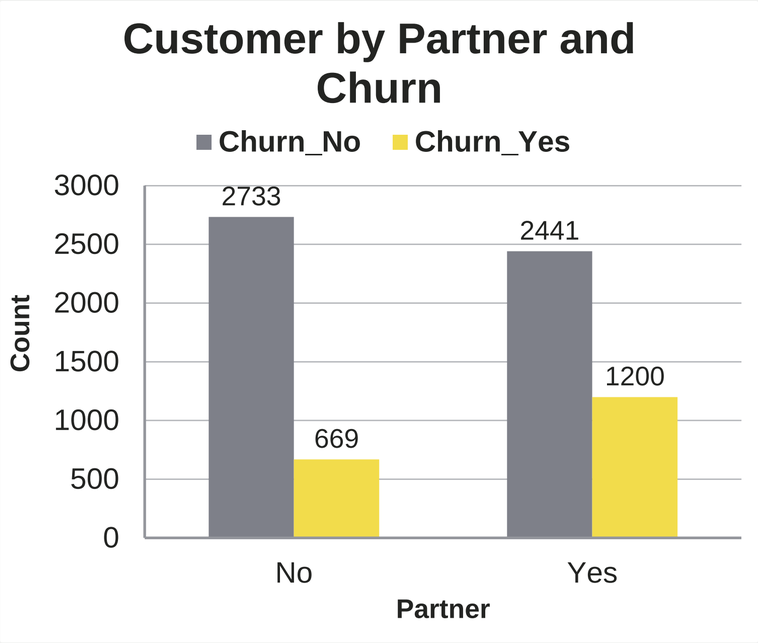



- Only 30% of the total customers have dependents.

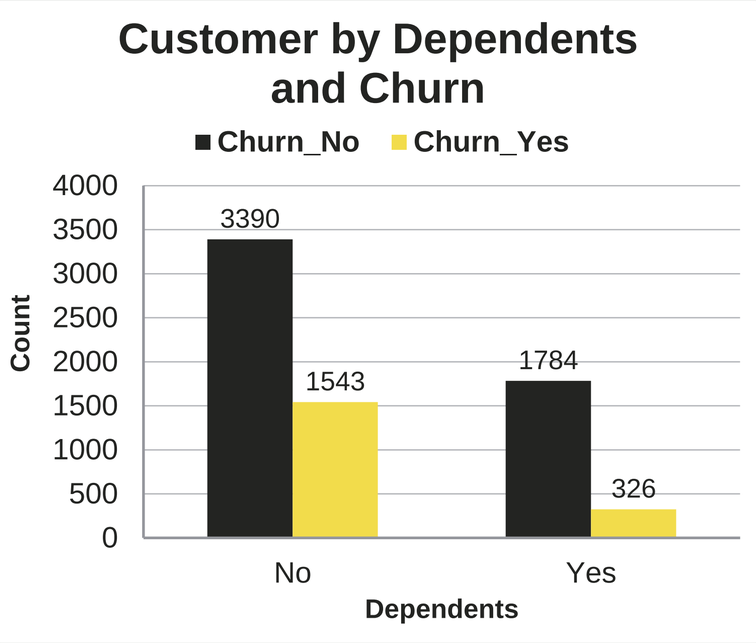
# Basic Information




- Customers who are senior citizen are **less likely to churn**.



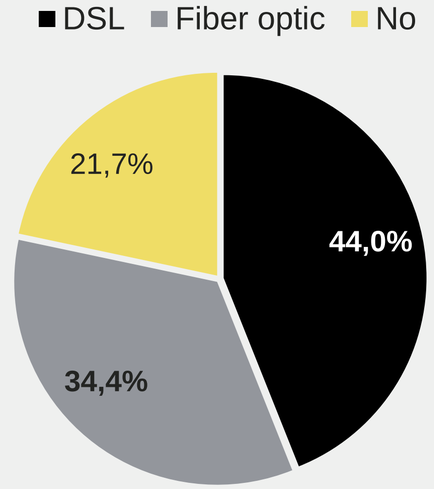

- Customers that have Partners are **more likely to churn.**




- Customers with Dependents are **less likely to churn.**

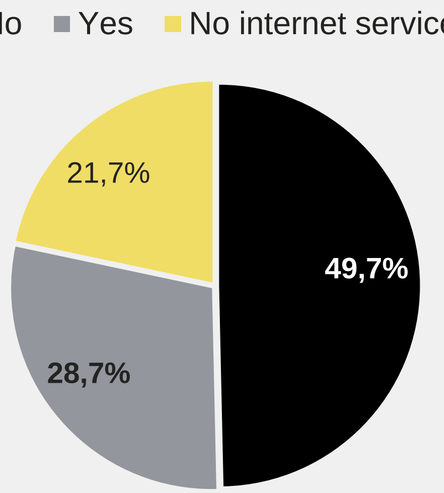
# Services Information

## Customer by Internet Service



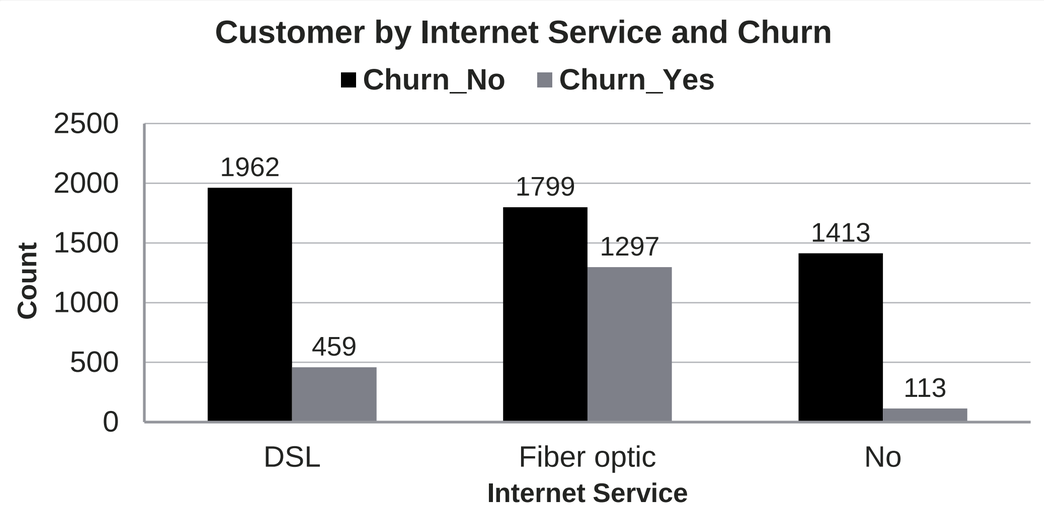

- Only 21,7% of the customers don’t have internet service.
- About 44% of the customers have internet services with a DSL connection and 34,4% with fiber optic.

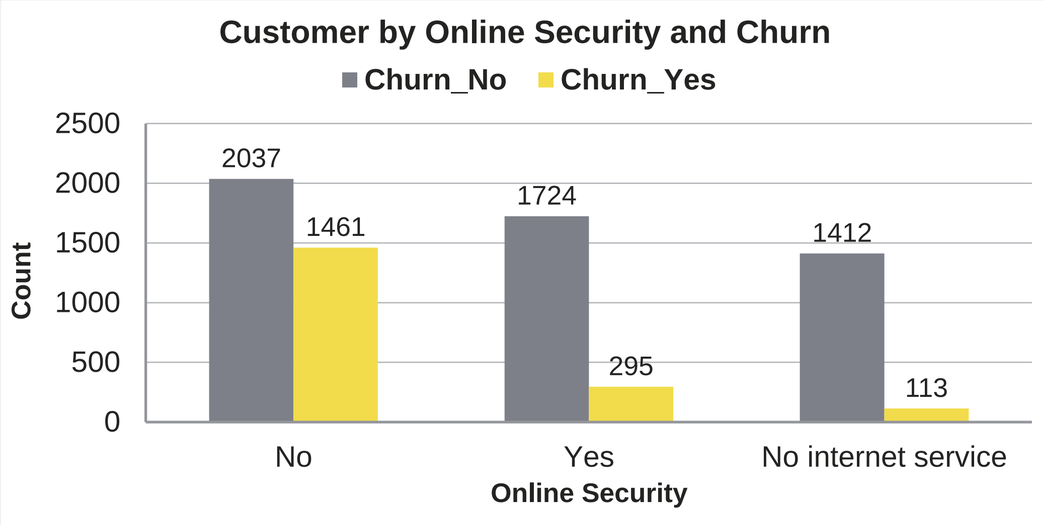
## Customer by Online Security




- About 28,7% of the customers don’t have online security.
- 49,7% of the customers have online security.

# Services Information

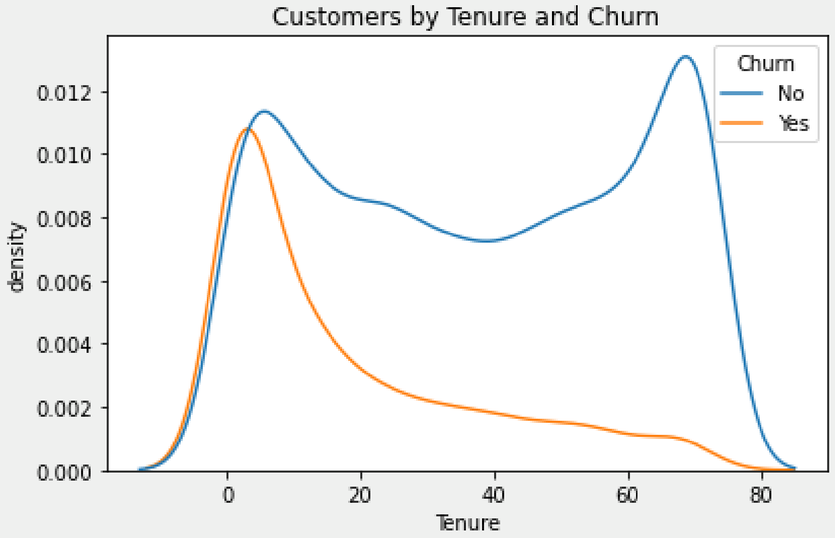


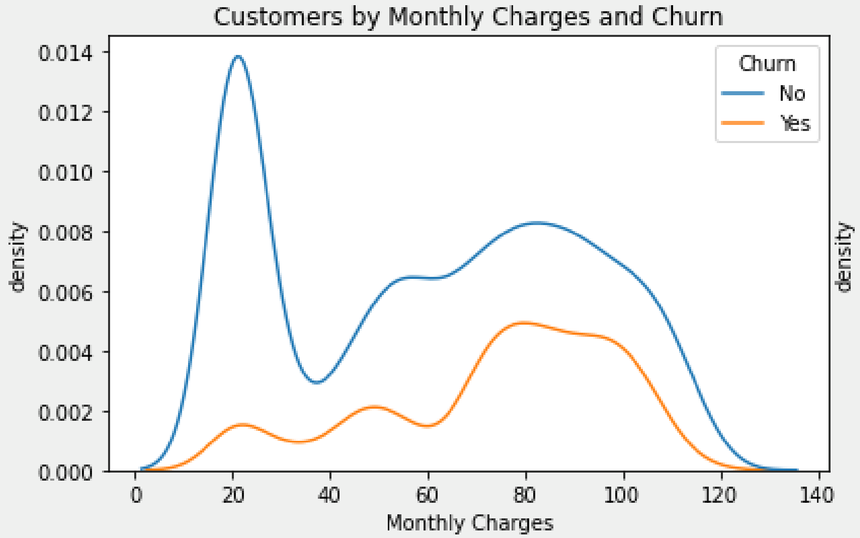



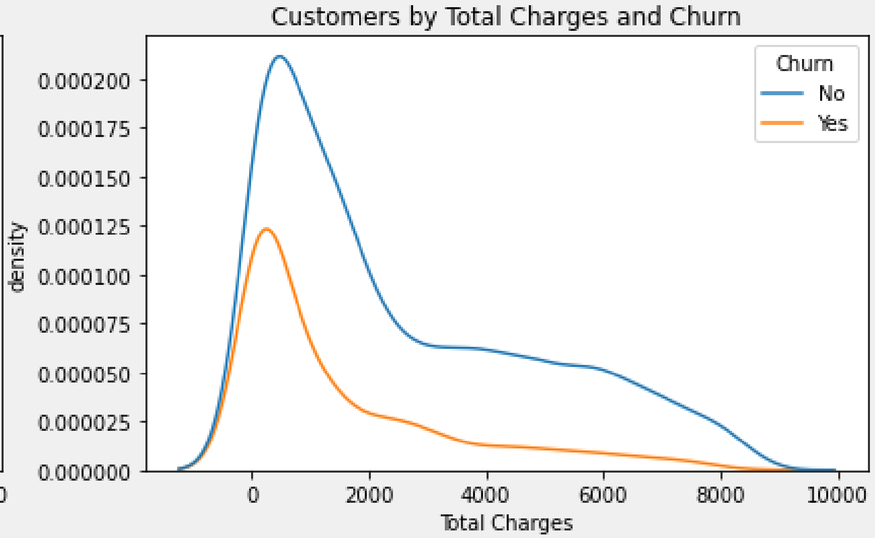

- Customers without internet service are **the least** count of customers.
- Customers with **Fiber optic are more probable to churn** than those with a DSL connection.

- Customers that don’t have online security are **more likely to churn.**

# Customer Account Information



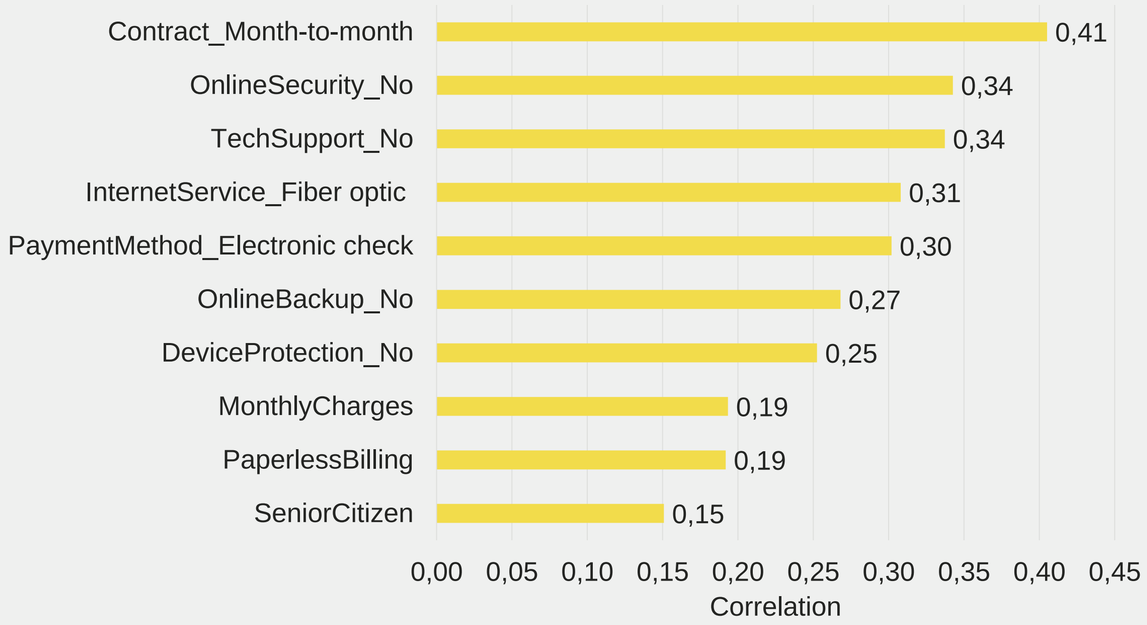



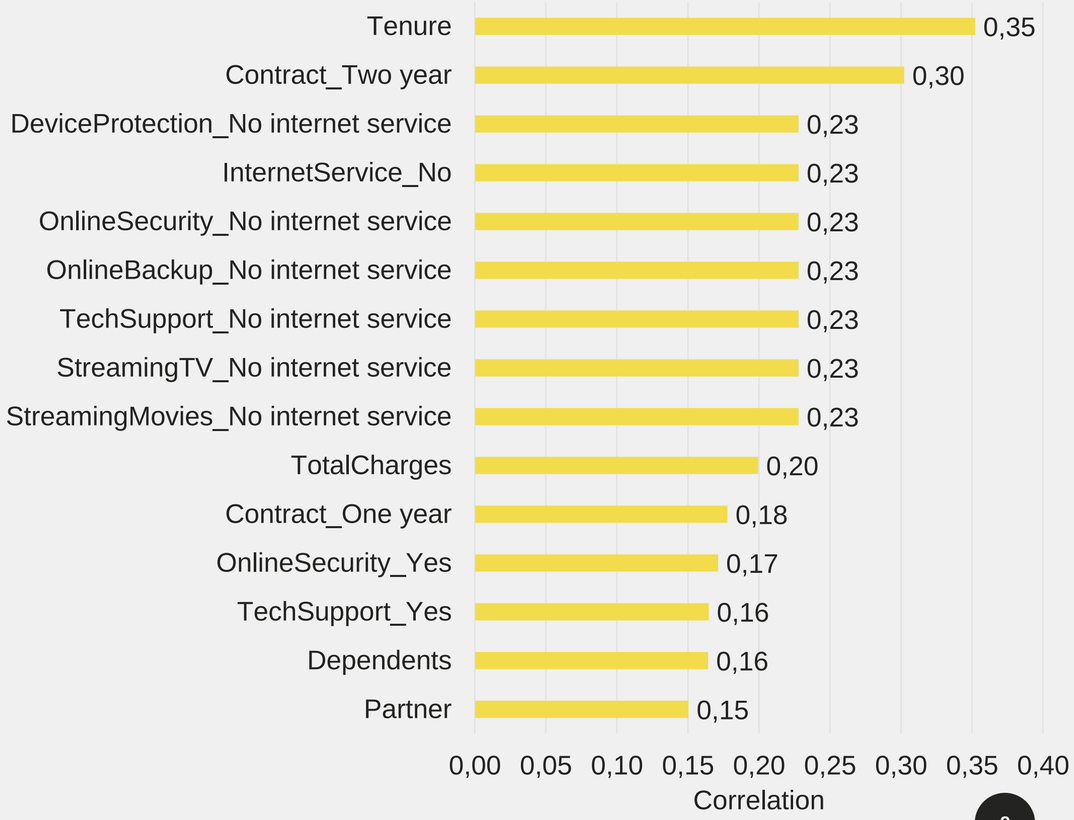

- **Recent costumers** are **more likely to churn.**
- Costumers with **higher Monthly Charges** are also **more likely to churn.**
- **Total Charges** don't have **difference between customer 'no-churn' and 'churn'.**

# Feature Selection

## Positive Correlation to Costumer Churn




## Negative Correlation to Costumer Churn




*correlation in absolute value

# Feature Selection

## Binary Column

- SeniorCitizen
- Partner
- Dependents
- PaperlessBilling

Encode : OrdinalEncoder()

## Categorical Column

- InternetService
- OnlineSecurity
- DeviceProtection
- TechSupport
- StreamingTV
- StreamingMovies
- Contract
- PaymentMethod

Encode: OneHotEncoder()

## Numerical Column

- Tenure
- MonthlyCharges
- TotalCharges

Scale: StandardScaler()

# Model Definition

## Model 1 ANN (Without Dropout, Sequential API)

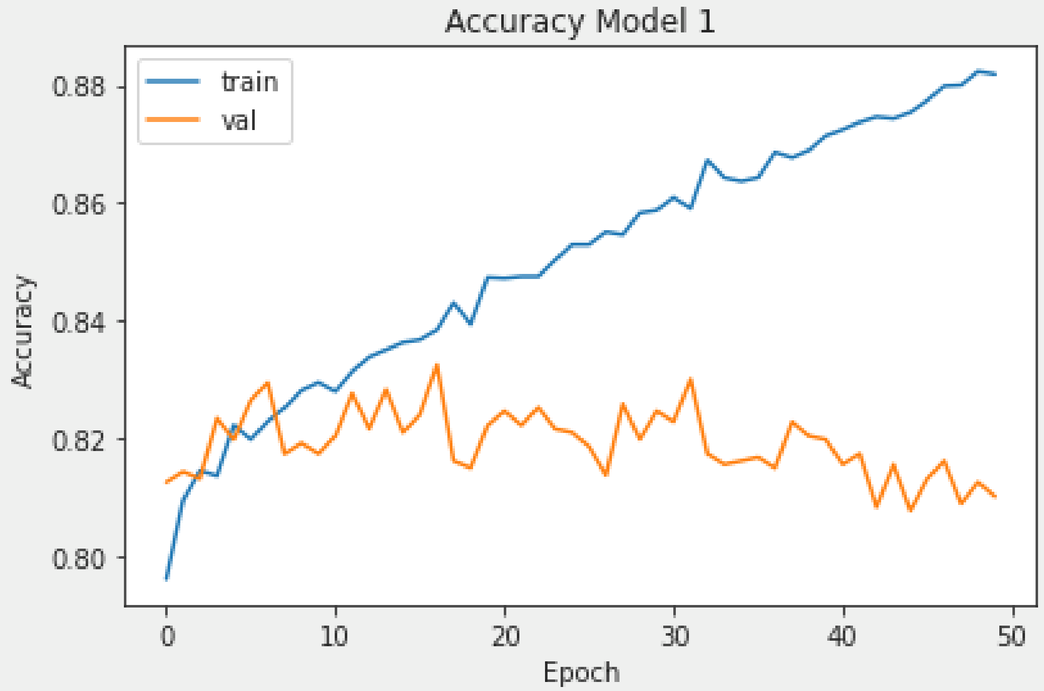
```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense (Dense)               (None, 64)                2304

 dense_1 (Dense)             (None, 32)                2080

 dense_2 (Dense)             (None, 1)                 33

=================================================================
Total params: 4,417
Trainable params: 4,417
Non-trainable params: 0
_________________________________________________________________
```

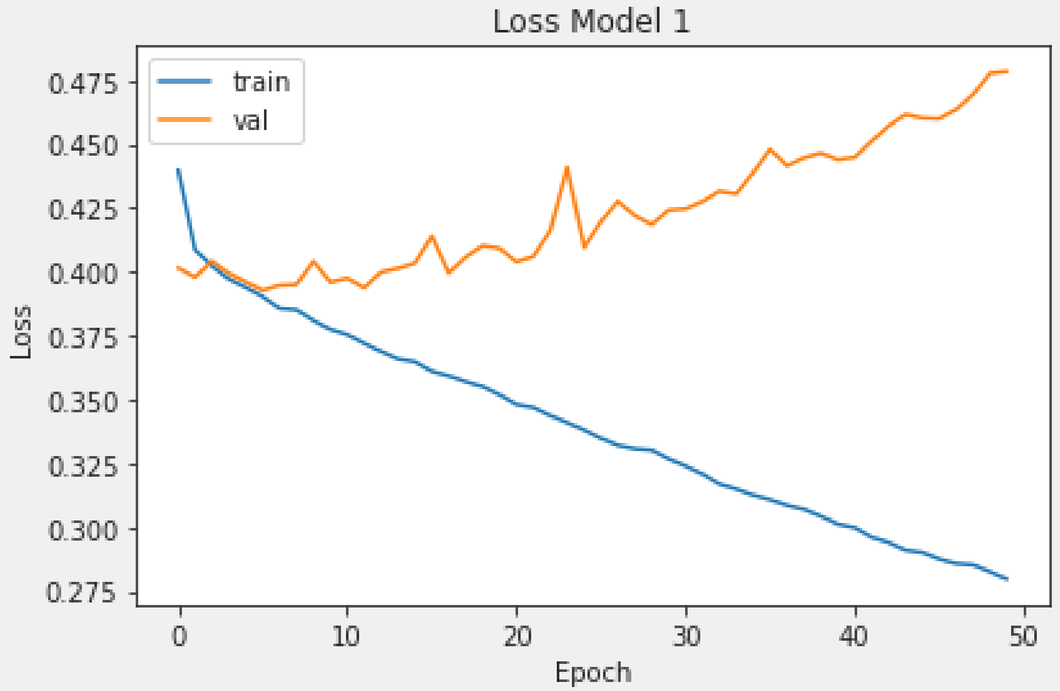
## Model 2 ANN (With Dropout 0.5, Functional API)

```
Model: "model"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 input_1 (InputLayer)        [(None, 35)]              0

 dense_3 (Dense)             (None, 64)                2304

 dropout (Dropout)           (None, 64)                0

 dense_4 (Dense)             (None, 32)                2080

 dropout_1 (Dropout)         (None, 32)                0

 dense_5 (Dense)             (None, 1)                 33

=================================================================
Total params: 4,417
Trainable params: 4,417
Non-trainable params: 0
_________________________________________________________________
```

# Model Evaluation

## Model 1 ANN (Without Dropout)

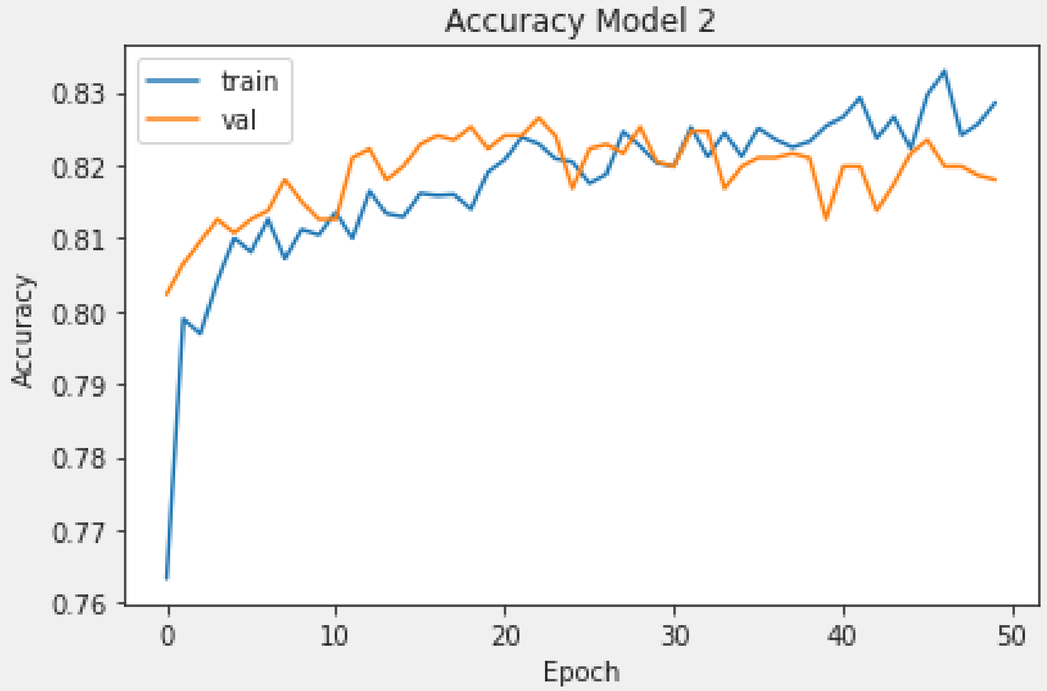


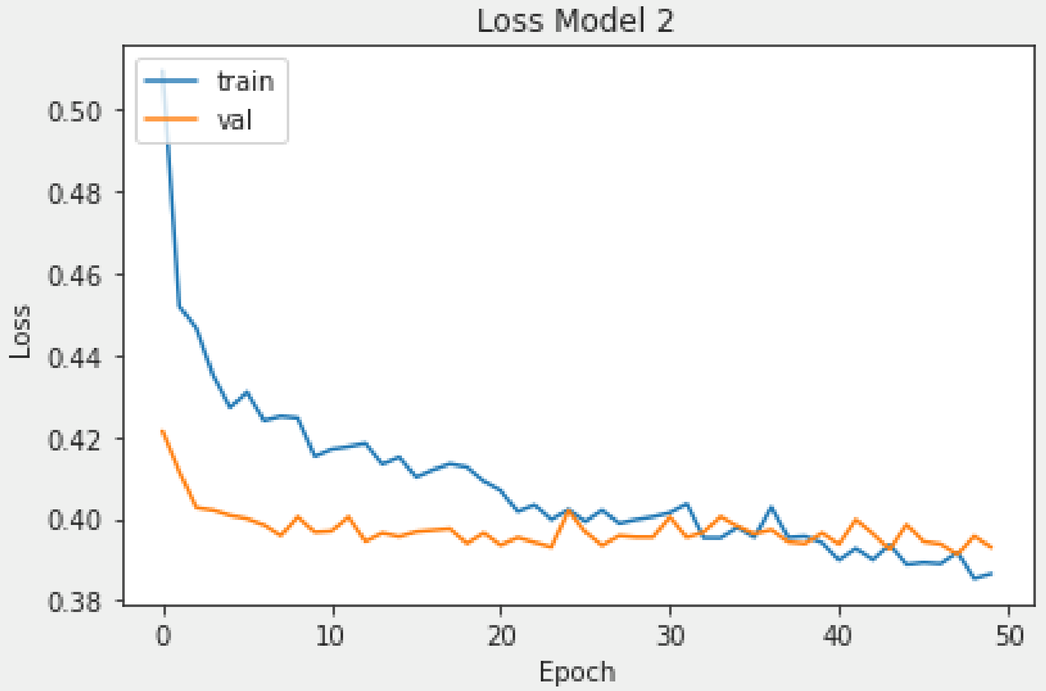




- **Model 1 is overfitting**

# Model Evaluation

## Model 2 ANN (With Dropout 0.5)




- **Model 2 is not overfitting (good fit)**

CUSTOMER CHURN PREDICTION

# Best Model Selection

| Model | Precision | Recall | F1-score | Accuracy |
|---|---|---|---|---|
| **Model 1 ANN (Without Dropout)** | 0.782727 | 0.841642 | 0.811116 | **0.806093** |
| **Model 2 ANN (With Dropout 0.5)** | 0.810891 | 0.800587 | 0.805706 | **0.808994** |

**Best Model : Model 2 ANN (With Dropout 0.5)**
Because Model 2 has the highest accuracy, and Model 2 is also not overfitting.

# Thank you!